

**Bibliographic data    Seed= JP2000250572**

Epoque PN	JP2000250572	20000914
Epoque AN	JP19990052173	19990301
Priority	JP19990052173	19990301

• Classifications:    IPC: G10L13/06; G10L13/08

    CI: G10L13/00

    AI: G10L13/06; G10L13/08

• Applicant:    NIPPON TELEGRAPH & TELEPHONE

• Inventors:    TANAKA KIMITO; MIZUNO HIDEYUKI; ABE MASANOBU; NAKAJIMA SHINYA

• Titles    :    TI: DEVICE AND METHOD FOR PREPARING VOICE DATABASE, DEVICE AND METHOD FOR PREPARING SINGING VOICE DATABASE

• Abstract PROBLEM TO BE SOLVED: To correctly and efficiently record a voice of required text, accent intonation, tempo and voice quality or a singing voice of required text, melody and voice quality. SOLUTION: The text, accent intonation, temp and voice quality corresponding to the required voice are inputted from a text input editor 1, a rhythm design editor 3 and a voice quality design editor 6, and a required synthesized voice is made by a reading giving part 2, a rhythm parameter generation part 4, a voice synthetic part 5 and a voice quality conversion part 7 to be outputted from a speaker 8, and that is let an utterer hear, and is let the same person imitate, and is let the same utter to be recorded from a micro-phone 9. Further, that is listened on the spot, and when that is the intended voice, by operating a switch 10 to be registered in a database 12.

**THIS PAGE BLANK (USPTO)**

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号  
特開2000-250572  
(P2000-250572A)

(43)公開日 平成12年9月14日(2000.9.14)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード(参考)
G 1 0 L	13/06	G 1 0 L	E 5 D 0 4 5
	13/08		H

審査請求 未請求 請求項の数 6 O L (全 6 頁)

(21)出願番号 特願平11-52173

(22)出願日 平成11年3月1日(1999.3.1)

(71)出願人 000004226

日本電信電話株式会社  
東京都千代田区大手町二丁目3番1号

(72)発明者 田中 公人

東京都新宿区西新宿3丁目19番2号 日本  
電信電話株式会社内

(72)発明者 水野 秀之

東京都新宿区西新宿3丁目19番2号 日本  
電信電話株式会社内

(74)代理人 100069981

弁理士 吉田 精孝

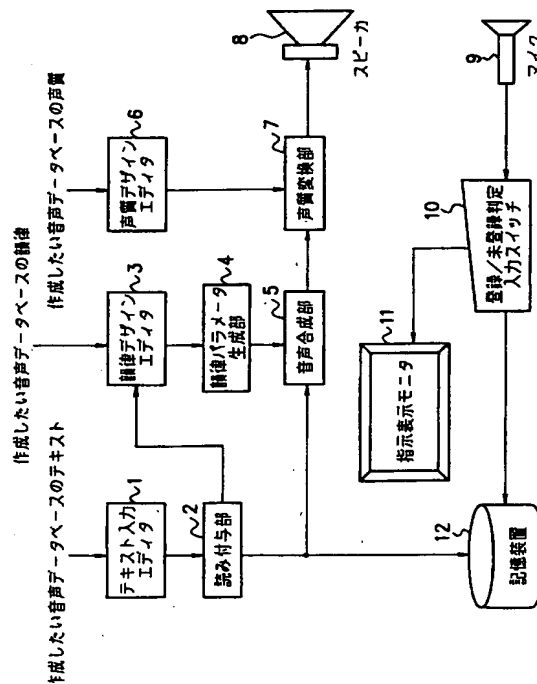
最終頁に続く

(54)【発明の名称】 音声データベース作成装置及びその方法並びに歌声データベース作成装置及びその方法

(57)【要約】

【課題】 所望のテキスト、アクセント・イントネーション、テンポ、声質の音声または所望の歌詞、旋律、声質の歌声を正確かつ効率的に収録可能とする。

【解決手段】 テキスト入力エディタ1、韻律デザインエディタ3及び声質デザインエディタ6から所望の音声に対応するテキスト、アクセント・イントネーション、テンポ及び声質を入力し、読み付与部2、韻律パラメータ生成部4、音声合成部5及び声質変換部7により所望の合成音声を作成してスピーカ8より出力し、これを発声者に聞かせ、真似させて発声させてマイク9から収録し、さらにこれをその場で聞いて意図通りの音声であればスイッチ10を操作することにより、データベース12に登録する。



## 【特許請求の範囲】

【請求項1】 人間が発声した音声を受録して、テキストとそれに対応する音声データからなる音声データベースを作成する音声データベース作成装置において、音声に対応するテキストを入力するテキスト入力部と、前記入力されたテキストの読み仮名を分析し、テキストに読み仮名を付与する読み付与部と、所望の韻律パターンを入力する韻律入力部と、前記韻律パターンから音声の韻律パラメータを生成する韻律パラメータ生成部と、前記読み仮名が付与されたテキストと前記韻律パラメータとに基づいて音声を合成する音声合成部と、所望の声質を示す情報を入力する声質入力部と、前記合成された音声を前記声質を示す情報に従って所望の声質に変換する声質変換部と、前記声質変換後の合成音声を再生する音声再生部と、前記再生された合成音声を真似た発声者による音声を収録する音声収録部と、前記収録された音声について前記音声データベースに登録するか否かの判断結果を入力する判断入力部と、前記判断結果を表示する指示表示部とを有し、前記収録された音声とそれに対応するテキストとを対応づけて前記音声データベースに登録することを特徴とする音声データベース作成装置。

【請求項2】 前記韻律入力部と前記声質入力部は各々韻律パターンと声質を示す情報を編集するための手段を有することを特徴とする請求項1記載の音声データベース作成装置。

【請求項3】 人間が発声した音声を受録して、テキストとそれに対応する音声データからなる音声データベースを作成する音声データベース作成方法において、音声に対応するテキストを入力し、前記入力されたテキストの読み仮名を分析し、テキストに読み仮名を付与し、所望の韻律パターンを入力し、前記韻律パターンから音声の韻律パラメータを生成し、前記読み仮名が付与されたテキストと前記韻律パラメータとに基づいて音声を合成し、所望の声質を示す情報を入力し、前記合成された音声を前記声質を示す情報に従って所望の声質に変換し、前記声質変換後の合成音声を再生し、前記再生された合成音声を真似た発声者による音声を収録し、前記収録された音声について前記音声データベースに登録するか否かを判断し、前記収録された音声とそれに対応するテキストとを対応づけて前記音声データベースに登録することを特徴とする音声データベース作成方法。

【請求項4】 人間が発声した歌声を受録して、歌詞と

それに対応する歌声データからなる歌声データベースを作成する歌声データベース作成装置において、歌声に対応する歌詞を入力する歌詞入力部と、前記入力された歌詞の読み仮名を分析し、歌詞に読み仮名を付与する読み付与部と、所望の旋律を入力する旋律入力部と、前記旋律から歌声の旋律パラメータを生成する旋律パラメータ生成部と、前記読み仮名が付与された歌詞と前記旋律パラメータとに基づいて歌声を合成する歌声合成部と、所望の声質を示す情報を入力する声質入力部と、前記合成された歌声を前記声質を示す情報に従って所望の声質に変換する声質変換部と、前記声質変換後の合成歌声を再生する歌声再生部と、前記再生された合成歌声を真似た歌唱者による歌声を収録する歌声収録部と、前記収録された歌声について前記歌声データベースに登録するか否かの判断結果を入力する判断入力部と、前記判断結果を表示する指示表示部とを有し、前記収録された歌声とそれに対応する歌詞とを対応づけて前記歌声データベースに登録することを特徴とする歌声データベース作成装置。

【請求項5】 前記旋律入力部と前記声質入力部は各々旋律と声質を示す情報を編集するための手段を有することを特徴とする請求項4記載の歌声データベース作成装置。

【請求項6】 人間が発声した歌声を受録して、歌詞とそれに対応する歌声データからなる歌声データベースを作成する歌声データベース作成方法において、歌声に対応する歌詞を入力し、前記入力された歌詞の読み仮名を分析し、歌詞に読み仮名を付与し、所望の旋律を入力し、前記旋律から歌声の旋律パラメータを生成し、前記読み仮名が付与された歌詞と前記旋律パラメータとに基づいて歌声を合成し、所望の声質を示す情報を入力し、前記合成された歌声を前記声質を示す情報に従って所望の声質に変換し、前記声質変換後の合成歌声を再生し、前記再生された合成歌声を真似た歌唱者による歌声を収録し、前記収録された歌声について前記歌声データベースに登録するか否かを判断し、前記収録された歌声とそれに対応する歌詞とを対応づけて前記歌声データベースに登録することを特徴とする歌声データベース作成方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、音声認識装置の学

習等に用いる音声サンプル、波形接続型の音声合成装置に用いる音声素片、音声応答装置で用いる応答音声、アニメーション・映画・マルチメディアコンテンツ作成のための音声または歌声のように、発声内容、アクセント・イントネーション、テンポ、声質等または歌詞、旋律、声質等が予め決まっている音声または歌声のデータベース作成作業において、作成者がイメージ・所望する音声または歌声を正確にかつ効率的に収録して音声または歌声データベースを構築するための音声データベース作成装置及びその方法並びに歌声データベース作成装置及びその方法に関するものである。

#### 【0002】

【従来の技術】従来の音声データベース作成装置は、発声者が発声する音声を収録するマイクと、その音声を記録する記録媒体とから構成されており、発声する内容はメモによって発声者に知らされ、韻律、漢字の読み、声質等は、データベース作成者が発声者に説明することによって伝えられていた。

【0003】また、従来の歌声データベース作成装置は、歌声を収録するマイクと、その歌声を記録する記録媒体とから構成されており、歌詞、旋律等はメモによって歌唱者に知らされ、歌詞の読み、声質等は、データベース作成者が歌唱者に説明することによって伝えられていた。

#### 【0004】

【発明が解決しようとする課題】しかし、このような従来の音声データベース作成装置では、登録したい音声の内容を記したメモを発声者に渡し、読み方、アクセント・イントネーション、テンポ、声質等について、データベース作成者が多大な時間をかけて説明した後、発声者はそれに従って発声する、というふうに多くの手間と人手と時間が必要であり、また、イントネーション、テンポ、声質等は発声者への説明が難しく、データベース作成者が所望する音声を収集することは非常に困難であった。

【0005】また、従来の歌声データベース作成装置では、登録したい歌声の歌詞や旋律を記したメモを歌唱者に渡し、声質等について、データベース作成者が多大な時間をかけて説明した後、歌唱者はそれに従って歌唱する、というふうに多くの手間と人手と時間が必要であり、また、声質等は発声者への説明が難しく、データベース作成者が所望する歌声を収集することは非常に困難であった。

【0006】本発明の目的は、データベース作成者が所望するテキスト、アクセント・イントネーション、テンポ、声質の音声を正確かつ効率的に収録して音声データベースに登録することができる音声データベース作成装置及びその方法、並びにデータベース作成者が所望する歌詞、旋律、声質の歌声を正確かつ効率的に収録して歌声データベースに登録することができる歌声データベー

ス作成装置及びその方法を提供することにある。

#### 【0007】

【課題を解決するための手段】前記目的を達成するため、本発明の音声データベース作成装置は、音声データベース作成者が所望する音声に対応するテキストを入力するテキスト入力部と、所望の韻律パターンを入力する韻律入力部と、所望の声質を示す情報を入力する声質入力部とを備え、さらにそれらの入力部から入力されたパラメータ通り正確に音声を合成する音声合成部を備えている。

【0008】実際の収録現場では、合成された音声を発声者が聞いた直後、それを真似て発声することにより、データベース作成者が所望するテキスト、アクセント・イントネーション、テンポ、声質の音声サンプルセットを正確かつ効率的に収録して音声データベースに登録することができる。

【0009】また、本発明の歌声データベース作成装置は、歌声データベース作成者が所望する歌声に対応する歌詞を入力する歌詞入力部と、所望の旋律を入力する旋律入力部と、所望の声質を示す情報を入力する声質入力部とを備え、さらにそれらの入力部から入力されたパラメータ通り正確に歌声を合成する歌声合成部を備えている。

【0010】実際の収録現場では、合成された歌声を歌唱者が聞いた直後、それを真似て歌唱することにより、データベース作成者が所望する歌詞、旋律、声質の歌声サンプルセットを正確かつ効率的に収録して歌声データベースに登録することができる。

#### 【0011】

【発明の実施の形態】次に、本発明について図面を参照して説明する。

【0012】図1は本発明の音声データベース作成装置の実施の形態の一例を示すもので、図中、1はテキスト入力エディタ、2は読み付与部、3は韻律デザインエディタ、4は韻律パラメータ生成部、5は音声合成部、6は声質デザインエディタ、7は声質変換部、8はスピーカ、9はマイク、10はスイッチ、11は指示表示モニタ、12は記憶装置である。

【0013】テキスト入力エディタ1は、音声データベース作成者が所望する音声に対応するテキストを入力するためのものである。読み付与部2は、前記入力されたテキストの発音（読み仮名）を分析し、前記テキストに分析した発音を対応づけて（読み仮名を付与して）出力する。

【0014】韻律デザインエディタ3は、所望するアクセント・イントネーションやテンポ等の韻律パターンをデザイン（編集）して入力する。韻律パラメータ生成部4は、入力された韻律パターンから音声の韻律パラメータを生成する。音声合成部5は、読み仮名が付与されたテキストと韻律パラメータとに基づいて音声を合成す

る。

【0015】声質デザインエディタ6は、所望する声質を示す情報をデザイン（編集）して入力する。声質変換部7は、前記合成された音声の前記声質を示す情報に従って所望の声質に変換する。

【0016】スピーカ8は、前記声質を変換した合成音声を出力する。マイク9は、スピーカ8から出力された合成音声を真似て発声者が発声した音声を収録する。

【0017】スイッチ10は、マイク9から収録された音声を音声データベース作成者が聞いてそれがデータベースに登録するのに十分なほど作成者の意図通り発声されているかどうかを判断した結果を入力するためのものである。指示表示モニタ11は、スイッチ10の出力を発声者に知らせる。

【0018】記憶装置12は、音声データベースを保持するもので、このデータベースにマイク9から収録された音声とそれに対応する（テキスト入力エディタ1から入力された）テキストを同時に登録する。

【0019】次に、前記装置を用いた音声データベース作成のようすをその動作とともに説明する。

【0020】データベース作成者は、収録作業の前に、本音声データベース作成装置に組み込まれているテキスト入力エディタ1を用いて所望の音声に対応するテキストを入力し、韻律デザインエディタ3を用いてそのテキストのアクセント・イントネーションやテンポ等の韻律パターンを所望するようにデザインして入力する。

【0021】韻律デザインエディタとしては、例えば、水野、阿部、中島、"音声表現力と制作効率を向上させた音声制作ツール Sesign98-" 日本音響学会平成10年度秋季研究発表会講演論文集、2-P-12、pp. 309-310、1998、10等に記載されたものがある。

【0022】さらにデータベース作成者は、声質デザインエディタ6を用いて所望する声質を示す情報をデザインして入力する。

【0023】前記入力されたテキストは読み付与部2にて読みが付与され、また、前記入力された韻律パターンは韻律パラメータ生成部4にて音声の韻律パラメータに変換された後、共に音声合成部5に入力されて合成音声となり、さらに声質変換部7にて前記入力された声質を示す情報に従う声質に変換される。

【0024】声質変換部（装置）としては、例えば、阿部、"波形処理による声質変換装置（VariVoice）" 日本音響学会平成9年度春季研究発表会講演論文集、3-7-6、pp. 269-270、1997、3等に記載されたものがある。

【0025】以上の処理により、音声データベース作成者は、自分がイメージ・所望する音声と同じ音響的性質を持つ合成音声を得ることができる。音声データベース作成者は、事前にそのような合成音声を作成しておく。

【0026】収録の現場では、その合成音声をスピーカ8から出力して発声者に順番に聞かせ、発声者にその合成音声を真似させて発声させ、これをマイク9から収録する。

【0027】その収録された音声を収録作業（データベース作成者と同じでも、そうでなくても良い。）が聞いて、直前の合成音声に音響的に近く、データベース作成者の意図に十分合っていると判断したならば、スイッチ10を操作する。これによってマイク9から収録された音声（前記入力されたテキストとともに記憶装置12（の音声データベース）に登録される。

【0028】なお、スイッチ10の操作出力は、指示表示モニタ11に表示され、発声者はその発声が十分であったかどうかを即座に知ることができる。

【0029】図2は本発明の歌声データベース作成装置の実施の形態の一例を示すもので、図中、21は歌詞入力エディタ、22は読み付与部、23は旋律デザインエディタ、24は旋律パラメータ生成部、25は歌声合成部、26は声質デザインエディタ、27は声質変換部、28はスピーカ、29はマイク、30はスイッチ、31は指示表示モニタ、32は記憶装置である。

【0030】歌詞入力エディタ21は、歌声データベース作成者が所望する歌声に対応する歌詞を入力するためのものである。読み付与部22は、前記入力された歌詞の発音（読み仮名）を分析し、前記歌詞に分析した発音を対応づけて（読み仮名を付与して）出力する。

【0031】旋律デザインエディタ23は、所望する旋律パターンをデザイン（編集）して入力する。旋律パラメータ生成部24は、入力された旋律パターンから歌声の旋律パラメータを生成する。歌声合成部25は、読み仮名が付与された歌詞と旋律パラメータとに基づいて歌声を合成する。

【0032】声質デザインエディタ26は、所望する声質を示す情報をデザイン（編集）して入力する。声質変換部27は、前記合成された歌声を前記声質を示す情報に従って所望の声質に変換する。

【0033】スピーカ28は、前記声質を変換した合成歌声を出力する。マイク29は、スピーカ28から出力された合成歌声を真似て歌唱者が歌唱した歌声を収録する。

【0034】スイッチ30は、マイク29から収録された歌声を歌声データベース作成者が聞いてそれがデータベースに登録するのに十分なほど作成者の意図通り歌唱されているかどうかを判断した結果を入力するためのものである。指示表示モニタ31は、スイッチ30の出力を歌唱者に知らせる。

【0035】記憶装置32は、歌声データベースを保持するもので、このデータベースにマイク29から収録された歌声とそれに対応する（歌詞入力エディタ21から入力された）歌詞を同時に登録する。

【0036】次に、前記装置を用いた歌声データベース作成のようすをその動作とともに説明する。

【0037】データベース作成者は、収録作業の前に、本歌声データベース作成装置に組み込まれている歌詞入力エディタ21を用いて所望の歌声に対応する歌詞を入力し、旋律デザインエディタ23を用いてその歌詞の旋律パターンを所望するようにデザインして入力する。

【0038】旋律デザインエディタとしては、例えば、水野、阿部、中嶋、"音声表現力と制作効率を向上させた音声制作ツール Sesign98" 日本音響学会平成10年度秋季研究発表会講演論文集、2-P-12、pp. 309-310、1998、10等に記載されたものがある。

【0039】さらにデータベース作成者は、声質デザインエディタ26を用いて所望する声質を示す情報をデザインして入力する。

【0040】前記入力された歌詞は読み付与部22にて読みが付与され、また、前記入力された旋律パターンは旋律パラメータ生成部24にて歌声の旋律パラメータに変換された後、共に歌声合成部25に入力されて合成歌声となり、さらに声質変換部27にて前記入力された声質を示す情報に従う声質に変換される。

【0041】声質変換部(装置)としては、例えば、阿部、"波形処理による声質変換装置(VarioVoice)" 日本音響学会平成9年度春季研究発表会講演論文集、3-7-6、pp. 269-270、1997、3等に記載されたものがある。

【0042】以上の処理により、歌声データベース作成者は、自分がイメージ・所望する歌声と同じ音響的性質を持つ合成歌声を得ることができる。歌声データベース作成者は、事前にそのような合成歌声を作成しておく。

【0043】収録の現場では、その合成歌声をスピーカ28から出力して歌唱者に順番に聞かせ、歌唱者にその合成歌声を真似させて歌唱させ、これをマイク29から収録する。

【0044】その収録された歌声を収録作業(データベース作成者と同じでも、そうでなくても良い。)が聞

いて、直前の合成歌声に音響的に近く、データベース作成者の意図に十分合っていると判断したならば、スイッチ30を操作する。これによってマイク29から収録された歌声が前記入力された歌詞とともに記憶装置32(の歌声データベース)に登録される。

【0045】なお、スイッチ30の操作出力は、指示表示モニタ31に表示され、歌唱者はその歌唱が十分であったかどうかを即座に知ることができる。

【0046】

10 【発明の効果】以上説明したように、本発明によれば、データベース作成者がイメージするアクセント・イントネーション、テンポ、声質で発声された音声サンプル、または所望する歌詞、旋律、声質で歌唱された歌声サンプルを、データベース作成者が所望する通り正確かつ短時間に収録して音声データベースまたは歌声データベースを構築することができる。さらに、データベース作成者は合成音声または合成歌声の作成さえしておけば、収録現場に作成者本人が立ち会うことなく代理人が収録作業を行うことも可能となる。

20 【0047】発声者または歌唱者にとっては、自分の発声または歌唱が十分データベース作成者の意図に合っていたかどうかを即座に知ることができるので、発声または歌唱の精度を向上させることができる。

【図面の簡単な説明】

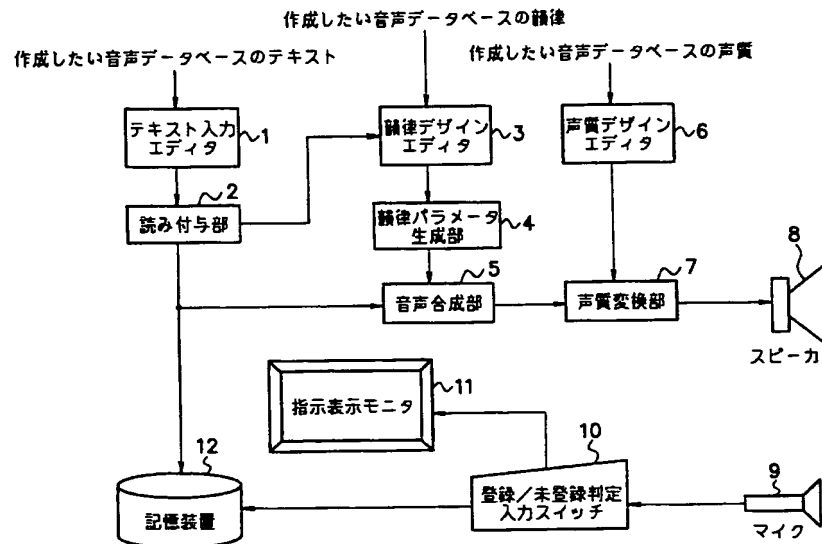
【図1】本発明の音声データベース作成装置の実施の形態の一例を示すブロック図

【図2】本発明の歌声データベース作成装置の実施の形態の一例を示すブロック図

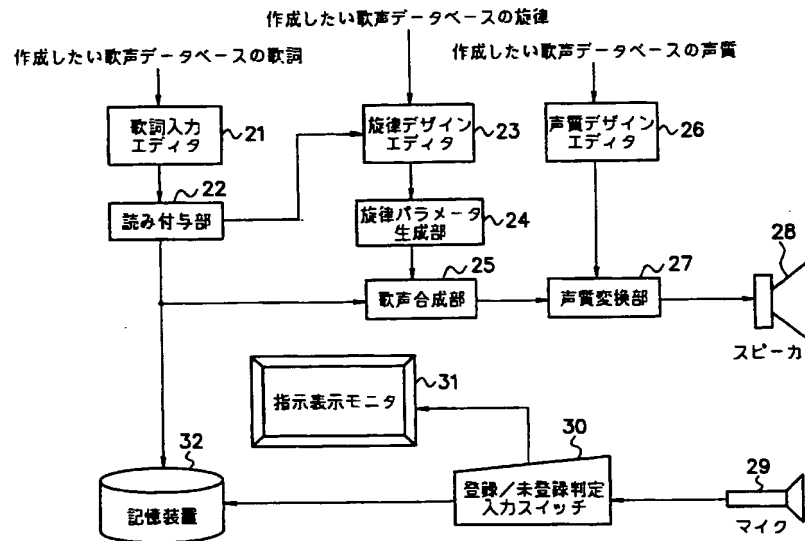
【符号の説明】

30 1: テキスト入力エディタ、2、22: 読み付与部、3: 韻律デザインエディタ、4: 韻律パラメータ生成部、5: 音声合成部、6、26: 声質デザインエディタ、7、27: 声質変換部、8、28: スピーカ、9、29: マイク、10、30: スイッチ、11、31: 指示表示モニタ、12、32: 記憶装置、21: 歌詞入力エディタ、23: 旋律デザインエディタ、24: 旋律パラメータ生成部、25: 歌声合成部。

【図1】



【図2】



フロントページの続き

(72)発明者 阿部 匡伸  
東京都新宿区西新宿3丁目19番2号 日本  
電信電話株式会社内

(72)発明者 中島 信弥  
東京都新宿区西新宿3丁目19番2号 日本  
電信電話株式会社内

Fターム(参考) 5D045 AA07 AA08 AA09

BEST AVAILABLE COPY